C-Statistic Fitting Routines

User's Manual and Reference Guide

John A. Nousek and Vida Farwana

Department of Astronomy and Astrophysics
The Pennsylvania State University
525 Davey Laboratory
University Park, PA 16802

Version v1.0

March 15, 1991

This program will read several input files and provide a best fit set of values for the function provided by the user, using either the C-statistic or the χ^2 statistic method. The program consists of one main routine and several functions and subroutines. Detailed description of each function and subroutine is covered in section 2.

A brief description of the C-statistic and the reason for its application is found in section 6.

Contents

1 Prior to Running the Program

The user must replace the function named 'PREDICT' with another function, which defines the equation that the user wishes to minimize. The sample function provided by the program minimizes the following equation: $f = N \cdot E^{-\gamma}$, where N= the normalization constant, E=energy in each bin and γ=slope. After providing the new function to the program, the user may or may not wish to modify the main program (cfit.f). In the main program all the input and output files are opened, input files are read, constant values are assigned and the 'POWELL' minimization subroutine is called 'NSRCH' times which will control the further flow of the program. 'NSRCH' is one of the assigned constants in the main program and will control how many times should the program try to minimize the equation. If NSRCH is 1, then; after the equation is minimized once, the program will terminate. If NSRCH is 2, the values calculated in the first round of minimization are used as the initial starting point and the equation is minimized again (NSRCH is set to 50 in the main program). Other constants assigned in the program which the user may wish to change, are:

EPSIL= 1.0E-20	Epsilon used as a default value for when the function becomes zero.
MODE = -1	Determines method of weighting least squares fit, used only when Chi-Square statistic minimization method is requested. if MODE=2 weight(i)=sigmay(i) (Pre-Calculated), where sigmay(i)= standard deviation of each observation. if MODE=1 weight(i)=1./(sigmay(i)**2) (Instrumental). if MODE=0 weight(i)=0. (No Weighting). if MODE=-1 weight(i)=1./y(i) (Statistical).
NFREE= 13	Number of degrees of freedom

Prior to execution of the program, the user must first compile the main program (cfit) along with the user's provided function (PREDICT) and all other supplied functions and subroutines; and then link them all together.

2 Functions and Subroutines

SLCTLU Selects the first available logical unit number for input and output files.

PREDICT function to be minimized.

In the sample program provided, the following function is minimized:

N*($E^{-\gamma}$) which is P(1)*XDATA**-P(2) in terms of the variables used in the program. As discussed earlier, the user must replace this function with another, which defines the equation of the function they wish to minimize. Following is the fortran program (for the above equation) used to test the program; make sure this routine is replaced with one which defines the function you wish to minimize.

```
C-------------------------------------------------------
C Name: PREDICT
C Filename: /usr/shue/cstat/predict.for
C Type: function
C
C Language: FORTRAN 77
C Purpose: to calculate the values in the power law
C by direct integration. function = N*E**-GAMMA
C N = Normalization Constant
C GAMMA = Slope
C E = Energy
C
C Subroutines: none
C
C Variables:
C
C Date: 1/19/88
C Author: David R. Shue
C
C-------------------------------------------------------


FUNCTION PREDICT (IVAL, P)

PARAMETER(NMAX=50, NBINS=100)
DIMENSION P(NMAX)

COMMON /CSTAT/ ISTAT, OBSERV(NBINS), XDATA(NBINS+1)

C---------
C calculate spectrum
C---------
AA = XDATA(IVAL)**(-P(2)+1.0)
BB = XDATA(IVAL+1)**(-P(2)+1.0)
DIFF = BB - AA
IF (P(2) .EQ. 1.) WRITE(*,*) 'divide by zero in predict ...'
TEMP = P(1)/(-P(2)+1.0)
PREDICT = TEMP * DIFF

C----
C end function
C----

RETURN
END
```

POWELL Minimization of a function FUNC of N variables. (FUNC is not an argument, it is a fixed function name). Input consists of an initial matrix XI whose dimensions are N by N, and whose columns contain the initial set of directions; and FTOL, the fractional tolerance in the function value such that failure to decrease by more than this amount on one iteration signals doneness. NPAR the number of parameters in the user's function. On output, P is set to the best point found, XI is then the current direction set, FRET is the returned function value at P, ITER is the number of iterations taken at P, and IERROR indicates if the number of iterations exceeded the maximum allowed (200).

USAGE: subroutine POWELL(P,XI,N,FTOL,ITER,FRET,IERROR)

LINMIN Linear minimization routine. Given an N dimensional point P and an N dimensional direction XI, moves and resets P to where the function FUNC(P) takes on a minimum along the direction XI from P, and replaces XI by the actual vector displacement that P was moved. Also returns FRET, the value of FUNC at the returned location P. This is all accomplished by calling the routines MNBRAK and BRENT.

USAGE: subroutine LINMIN(P, XIN, N, FRET)

MNBRAK Given a function F1DIM, and given distinct initial points AX and BX, this routine searches in the downhill direction (defined by the function as evaluated at the initial points) and returns new points AX, BX and CX which bracked a minimum of the function. Also returned are the function values at the three points FA,FB and FC.

USAGE: subroutine MNBRAK(AX, BX, CX, FA, FB, FC, F1DIM)

BRENT Given a function F1DIM, and given a bracketing triplet of abcissas AX, BX and CX (such that BX is between AX and CX, and F(BX) is less than both F(AX) and F(CX)), this routine isolates the minimum to a fractional precision of about TOL using BRENT's method. The abcissa of the minimum is returned as XMIN, nd the minimum function value is returned as BRENT, the returned function value.

USAGE: function BRENT(AX, BX, CX, F1DIM, TOL, XMIN)

F1DIM Accompanies LINMIN.

Constructed by LINMIN, F1DIM is the value of FUNC along the line through the point P in the direction XI. F1DIM is an artificial function of one variable, which is the value of the function FUNC. LINMIN communicates with F1DIM through a common block, it then calls our familiar one-dimensional routines MNBRAK and BRENT and instructs them to minimize F1DIM.

USAGE: function F1DIM(ANUMBER)

FUNC Evaluate C-statistic for fit to data

C = 2 * $\sum$(YFIT - Y * ln(YFIT))

OR ....

Evaluate reduced chi-square for fit to data

FCHISQ = $\sum$((Y - YFIT)**2 / SIGMA **2) / NFREE

where:

YFIT = PREDICT(I,P) = value of the function PREDICT

3 Input and Output Files

For description of the variable names refer to section 4.

RUNFIT.DAT (Input file)
line 1, ISTAT and FTOL
line 2, P(1) and XI(1,J) for j=1 .. NPAR
repeat line 2 for each of the parameters in the function defined in PREDICT.

OBSERV.DAT (Input file)
line 1, OBSERV(1)
repeat line 1 for each bin.

XDATA.DAT (Input file)
line 1, XDATA(1)
repeat line 1 for energy in each bin.

SIGMAY.DAT (Input file)
line 1, SIGMAY(1)
Repeat line 1 for standard deviation of each observ(I).
NOTE: This file is used only if the chi-square minimization is selected and the desired MODE. The default value of -1 (statistical method) for MODE is hard coded in the main program and must be changed if other methods are preferred. (refer to section 4 for possible options for MODE).

RUN.DAT (Output file)
This file contains NSRCH+1 lines, where NSRCH is the number of times that function should be minimized. It is assigned in the main program.
first line indicates the minimization method (ISTAT) selected 1=C-statistic and 2=Chi-square.

second through last line contains the following information:
- iteration number
- P(1) through P(I), where I is the number of parameters in the function PREDICT
- FRET
- ITER

4 Variable Description

ISTAT	= 1	Minimize using the C-statistic method
	= 2	Minimize using the Chi-square statistic method
FTOL		The criterion for doness (real). Fractional tolerance in the function value such that failure to decrease by more than this amount on one iteration signals doneness.
MODE		Method of weighting least squares fit (used when ISTAT = 2)
	= 2	(precalculated) weight(i) = sigmay (i)
	= 1	(instrumental) weight(i) = 1./(sigmay(i)**2)
	= 0	(no weighting) weight(i) = 0.
	=-1	(statistical) weight(i) = 1./y(i)

NFREE	Number of degrees of freedom
NPAR	Number of parameters in the user's function
ITER	Number of iterations (integer)
FRET	Returned function value at P (real)
P(I)	Initial point for each parameter in function PREDICT for I=1 .. NPAR
XI(I,J)	New set of direction for each parameter I for I=1 .. NPAR and J=1 .. NPAR
OBSERV(I)	Array representing the number of observation in each bin. (I=1 .. NBINS)
XDATA(I)	Energy in each bin, (I=1..NBINS)
SIGMAY(I)	Standard deviations of each observation (used when ISTAT=2 and MODE > 0) for I=1 .. NBINS

5 Sample Input and Output Files

In this section we provide an example of input files and the files returned as output. These can be used as an example and as a test case to verify your software installation. Note that the input and output files listed here are for the program running with the test function 'predict' defined in section 2.

- RUNFIT.DAT (Input file)

```
1,.0001
1.0,1.0,1.0
.50,1.0,.50
```

- OBSERV.DAT (Input file)

```
324.000
252.000
205.000
199.000
197.000
175.000
157.000
136.000
138.000
136.000
129.000
111.000
130.000
125.000
117.000
```

- XDATA.DAT (Input file)

```
0.095
0.145
0.195
0.245
0.295
0.345
0.395
0.445
0.495
0.545
0.595
0.645
0.695
0.745
0.795
0.845
```

- SIGMAY.DAT (Input file)

```
1.41
0.00
1.41
1.41
2.00
1.73
0.00
1.00
0.00
1.73
0.00
1.41
1.00
0.00
1.41
```

- RUN.DAT (Output file)

```
1
1 2016.20 0.524285 -21147.2 12
2 2016.20 0.525515 -21147.2 1
3 2016.20 0.525515 -21147.2 1
```

6 Discussion of the C-Statistic

In X-ray astronomy, the most common technique to fit data to models has been the application of a non-linear regression to a hypothetical source spectrum convolved with the response of the detector, seeking to minimize the χ^2 value generated by the model and the data set. This basic approach was first applied in X-ray astronomy by Gorenstein, Gursky and Garimire (1968).

Two advantages to this technique are, first, that the fitting procedure is 'robust', meaning that the fit will generally converge, and secondly, that the technique is 'efficient', meaning that the confidence intervals on the parameters and the probabilities for rejecting the model are as strong as possible. Confidence intervals define the range allowed on each parameter within the significance level we choose (for example, random noise variations will allow the best fit parameter to vary within the 90% confidence interval 90% of the time).

Despite the generally satisfactory results of using this technique, we run into problems using it with small numbers of events. An important mathematical advantage of χ^2 fitting is that, for adequately large samples, χ^2 fitting is asymptotically independent of the shape (i.e. distribution) of the events in the sampling bins, under the null hypothesis (Lindgren 1976, for example). (The null hypothesis is the assumption that the only deviation between model and data is random measurement error.) Unfortunately, in many cases in astronomy, and in X-ray astronomy in particular, the data samples are not large enough to assure χ^2 reaches the asymptotic limit. In practical terms the important question is when and how severely χ^2 fitting departs from its asymptotic behavior in the context of X-ray astronomy.

Nousek and Shue (1989) discuss this problem. In this work ideal models having a power law distribution of photons ($N(E) = N_o \cdot E^{-\gamma}$), were used to generate data sets. The data sets differed from the ideal by being drawn randomly, bin by bin, from a Poisson distribution having a mean equal to the ideal mean. Thus the data sets correspond to actual samplings of the ideal distribution under the conditions of statistical fluctuations that we expect in a photon counting experiment. By applying a fitting procedure to the generated data sets we attempt to reconstruct the parameters (N_o and γ in this case) that were used to generate the data.

If the fitting is unbiased the parameters returned as the best fit should approximate the original ones. The degree to which they do not should be a reflection of the noise introduced by the statistical fluctuations, and should be consistent with the confidence interval on the parameters predicted by the fitting. By repeating the process with additional random data sets the fluctuations should average to zero, leading to a more precise reconstruction of the initial parameters. Thus by accumulating the mean of the best fit parameters we should find it tending to the original value.

The result of our work for the simple χ^2 fitting is that χ^2 does not reconstruct the initial values when the number of counts per bin is low. The following table illustrates the systematic bias introduced into the best fit parameter by using χ^2 fitting when few events are found in some bins. The table contains the ratio of the mean best fit parameters to the true value used to generate the data sets. Hence if the fitting supplied no bias the ratios should approach one. Instead the ratios show clear and significant deviations from one. (Note only fits that converged are included in the tabulated means.)

Table I. Chi-Squared Minimization - Marquardt's Method (250 Fits) Results for Parameters and Convergence						
	Wide Boundaries			Narrow Boundaries		
N	N_{calc}/N_o	γ_{calc}/γ	Percent Converg.	N_{calc}/N_o	γ_{calc}/γ	Percent Converg.
25	.411	1.455	.74	.611	1.180	1.00
50	.457	1.365	.73	.633	1.141	1.00
75	.522	1.213	.79	.639	1.129	1.00
100	.558	1.156	.76	.647	1.119	1.00
150	.655	1.095	.81	.723	1.092	1.00
250	.713	1.086	.89	.764	1.072	1.00
500	.820	1.052	.96	.855	1.042	1.00
750	.876	1.032	.98	.904	1.026	1.00
1000	.916	1.020	1.00	.943	1.016	1.00
2500	.958	1.010	1.00	.977	1.006	.98
5000	.976	1.004	1.00	.982	1.005	.93
10000	.985	1.003	1.00	.990	1.002	.88

The χ^2 fitting used Marquardt's method for finding the best fit, with additional constraints limiting the range allowed for the parameters. In the table above the 'Wide Boundaries' case set these constraints so loosely that they did not restrict the fitting procedure, while in the 'Narrow Boundaries' case the constraints kept the final result relatively close to the ideal values. The two cases simulate the effect of a totally unconstrained 'automatic' fitting (the 'Wide' case), and a carefully hand selected 'manual' fitting (the 'Narrow' case).

At first glance it would appear that careful constraints applied as in the 'Narrow' case would surmount the small count problem. The bias is smaller and the fits all converge. The next table reveals the difficulty.

Table II. χ^2 Minimization - Marquardt's Method Results for Minimum χ^2 and σ				
	Wide Boundaries		Narrow Boundaries	
N	Reduced χ^2	σ	Reduced χ^2	σ
25	.714	1.801	.677	.546
50	.866	.955	.785	.370
75	.887	.610	.846	.260
100	.988	.394	.912	.221
150	1.069	.328	.946	.192
250	1.151	.232	1.038	.143
500	1.160	.192	1.030	.104
750	1.120	.170	1.030	.080
1000	1.088	.067	1.025	.068
2500	1.132	.055	.982	.038
5000	1.142	.027	.999	.026
10000	1.057	.019	.985	.019

The Reduced χ^2 is the mean of the convergent best fit value, and the σ is the r.m.s. of the distribution of best fit value for γ. At large N both cases have similar σ. At small N the 'Narrow'

case the σ becomes comparable to the width of the boundaries (three in this case) and the apparent accuracy of the fit comes purely from choosing the boundaries symmetrically about the ideal answer. (Note also that the violation of the assumption of Gaussian statistics causes the Reduced χ^2 to fall below one.)

In short traditional χ^2 fitting techniques can lead to biases of 50% for very small count data sets, and errors of 10% even if 100-1000 counts are available. Efforts to improve the fitting by tight parameter constraints result in achieving fits which only reflect the prejudices of the constraints.

The issue of low count statistics in photon counting experiments has been directly addressed in the work of Cash (1979). He proposes a maximum likelihood statistic based on the Poisson, rather than Gaussian probability distribution. His statistic, called the C-statistic, is defined,

$$C = 2\sum_{i=1}^{N}(m_i - n_i \ln(m_i))$$

and m_i and n_i are the number of counts predicted by the model and observed, respectively. Operationally this proceeds exactly as in the χ^2 minimization case, including the inference of confidence intervals on the parameters. (For one parameter the 68% confidence interval is found by finding the parameter values where $C = C_{min} + 1$. For more parameters, this becomes $C = C_{min} + f(\nu)$ where $f(\nu)$ can be found in a table in Lampton, Margon and Bowyer 1976).

Cash shows that the C-statistic goes to χ^2 in the limit of large n, and that it should be more efficient than χ^2. He also explicitly applies it to the case where each photon is alone (i.e. only one or zero photons in every bin).

Table III lists the results of using χ^2 and the C-statistics. The inital starting vectors of the fits for the two statistics were the same.

Table III. Comparison of χ^2 and C-Statistic (250 Fits)

	χ^2 Minimization			C-Statistic Minimization		
N	N_{calc}/N_o	γ_{calc}/γ	Fraction Converged	N_{calc}/N_o	γ_{calc}/γ	Fraction Converged
25	.709	1.152	.96	1.269	.958	.86
50	.647	1.134	1.00	1.079	.998	1.00
75	.636	1.130	1.00	1.078	.995	1.00
100	.673	1.109	1.00	1.053	.996	1.00
150	.707	1.094	1.00	1.015	1.005	1.00
250	.767	1.072	1.00	1.019	1.000	1.00
500	.863	1.040	1.00	.997	1.004	1.00
750	.905	1.025	1.00	.997	1.002	1.00
1000	.937	1.017	1.00	1.001	.999	1.00
2500	.973	1.007	1.00	1.005	1.000	1.00
5000	.988	1.003	1.00	.984	1.003	1.00
1000	.996	1.001	1.00	1.003	1.000	1.00

The immediate conclusion is that for this problem the Cash statistic introduces virtually no systematic bias to the fit results.

The principal disadvantage of the C-statistic is that there is no value corresponding to the Reduced χ^2 value with which we can measure the goodness of the fit. For the C-statistic there exist no analogous tables with which the goodness of fit can be determined. We can only determine the best parameters by minimizing the function, but we have no criteria to reject the model.

The reason general tables can not be produced is that the distribution in each bin is different, and depends on the model. This exactly the regime in which χ^2 fails. One could simulate the effect of random deviations for the given model and observed data set, but the simulation would have to be repeated for each data set and each model.

An alternative approach to low count statistics of relatively recent development in modern statistical theory is the bootstrap. Rather than assuming knowledge of the distribution function the bootstrap uses the observed data as an empirical sample of the distribution. Analysis proceeds by generating randomized data sets from the observed one, and fitting the randomized ones using any proper fitting technique. The distribution of fit results should tend to the true one, and the uncertainty can be estimated from the width of the distribution.

The key to the bootstrap is the generation of randomized data sets. The simplest visualization in the case of X-ray astronomy is to imagine that a marble corresponding to each observed photon is placed in a bag; the bag is shaken and a marble chosen at random. After recording which marble was found the marble is returned to the bag and the process is repeated until a number of events equal to the observed total has been collected. The resulting data set is not identical to that observed, but it has a very similar statistical nature, within the fluctuations expected from random chance. Hence the result of repeating the fitting on the similar data sets is to give a very good estimate in the uncertainty without having to assume knowledge of the underlying probability distribution.

The bootstrap is a subject of active statistical research (Efron 1982), and holds great potential, particularly for its ability to accurately gauge the confidence intervals on the parameters.

7 Acknowledgements

Our colleagues, Mr. David Shue and Mr. Chris Frye, have made valuable contributions to writing and testing many of the routines described in this manual. Funding for the preparation of this manual and software were provided by NASA's Science Operations Branch under NASA Grant No. NAG5-1211.

8 References

1. Cash, W. 1979, *Ap. J.*,**228**, 939.

2. Efron, Bradley, 1982, "The Jackknife, the Bootstrap and Other Resampling Plans," CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 38.

3. Gorenstein, P., Gursky, H., and Garmire, G. 1968, *Ap. J.*,**153**, 885.

4. Lindgren, Bernard W., 1976, "Statistical Theory," (University of Minnesota, Third Edition, Macmillian), 424.

5. Nousek, J. A., and Shue, D. R. 1989, *Ap. J.*, **342**, 1207.